



Pengelompokan Kemiskinan di Provinsi Jawa Timur Tahun 2024 Dengan Pendekatan Algoritma K-Means Dan K-Means++

Suci Amalia^{1*}, Annisatul Nikmah², Lutfiah Maharani Siniwi³

¹STIKES Sukma Wijaya, Indonesia,

²Universitas Sebelas Maret, Indonesia,

³Universitas Jenderal Soedirman, Indonesia

Email: amaliasuci117@gmail.com*, annisatulnikmah@staff.uns.ac.id, lutfiah.maharani@unsoed.ac.id

Manuscript accepted:

Revised:

Date of publication:

KATA KUNCI

Kemiskinan, Clustering, K-Means, K-Means++, Jawa Timur.

ABSTRAK

Kemiskinan merupakan masalah kompleks yang mencerminkan kesejahteraan masyarakat dan kinerja pembangunan suatu daerah. Di Provinsi Jawa Timur, walaupun persentase kemiskinan mengalami penurunan hingga mencapai 9,79% pada tahun 2024, namun kesenjangan antarwilayah masih menjadi tantangan. Penelitian ini bertujuan untuk mengetahui karakteristik dan mengelompokkan Kabupaten/Kota di Jawa Timur berdasarkan indikator kemiskinan, yakni jumlah penduduk miskin, tingkat pengangguran terbuka (TPT), harapan lama sekolah (HLS), rata-rata lama sekolah (RLS), tingkat partisipasi angkatan kerja (TPAK), dan gini ratio. Algoritma clustering yang digunakan adalah K-Means dan K-Means++, untuk membandingkan hasil dan performa keduanya. Penentuan jumlah cluster optimum didasarkan pada nilai Davies-Bouldin Index (DBI) dan Silhouette Score, untuk evaluasi performa algoritma dilihat melalui nilai inertia. Hasil penelitian menunjukkan bahwa jumlah cluster optimum yang terbentuk adalah tiga cluster. Cluster 1 terdiri dari wilayah dengan kemiskinan struktural, Cluster 2 terdiri dari wilayah dengan kemiskinan fungsional, dan Cluster 3 mencerminkan wilayah dengan kondisi indikator yang relatif sedang. Berdasarkan hasil evaluasi, algoritma K-Means++ menghasilkan nilai inertia yang lebih rendah dibandingkan K-Means, serta nilai DBI dan Silhouette yang lebih baik, sehingga dapat disimpulkan bahwa K-Means++ lebih optimal. Hasil penelitian ini diharapkan dapat menjadi acuan dalam merancang strategi kebijakan penanggulangan kemiskinan berbasis karakteristik wilayah yang lebih tepat sasaran.

PENDAHULUAN

Kemiskinan merupakan permasalahan yang kompleks dan merepresentasikan tingkat kesejahteraan masyarakat serta kinerja pembangunan dan pertumbuhan ekonomi; karena itu, penanggulangannya menjadi fokus utama pembangunan di Indonesia, termasuk di Provinsi Jawa Timur (World Bank, 2023; UNDP, 2023; Badan Pusat Statistik, 2024; Badan Pusat Statistik Provinsi Jawa Timur, 2025; Suryahadi, Izzati, & Suryadarma, 2020).

Angka kemiskinan di Provinsi Jawa Timur cenderung mengalami penurunan dari tahun ke tahun, di mana tahun 2020 persentase kemiskinan mencapai 11,09% dan mengalami kenaikan di tahun berikutnya. Namun, pada tahun 2022 persentase kemiskinan kembali turun menjadi 10,38% dan terus mengalami penurunan hingga mencapai angka 9,56% pada September tahun 2024 (Bappeda Provinsi Jawa Timur, 2025). Penurunan tersebut menunjukkan bahwa adanya upaya perbaikan kemiskinan di Jawa Timur. Meskipun begitu, pemberantasan kemiskinan harus dioptimalkan, mengingat kesenjangan antar wilayah masih ada.

Faktor penyebab kemiskinan bersifat multidimensional. Di Indonesia, tingkat pengangguran terbuka berkorelasi positif dengan kemiskinan, sementara modal manusia—yang tercermin dari harapan lama sekolah dan rata-rata lama sekolah—berkorelasi negatif karena peningkatan pendidikan memperbaiki produktivitas dan peluang kerja; selain itu, capaian pembangunan manusia (IPM/HDI) secara konsisten terkait dengan penurunan kemiskinan dalam analisis panel terbaru lintas provinsi. Bukti empiris mutakhir mendukung keterkaitan tersebut: dinamika kemiskinan dipengaruhi berbagai determinan struktural (Purwono, Wardana, Haryanto, & Mubin, 2021); pengangguran memperburuk kemiskinan perkotaan (Faturohim, Akbar, Hidayat, & Saksono, 2023); rata-rata lama sekolah berpengaruh terhadap persentase kemiskinan (Chamidah, Mardianto, Limanta, & Hastuti, 2020); pendidikan dan modal manusia menurunkan risiko kemiskinan (Taufiq & Dartanto, 2020); dan IPM berasosiasi negatif dengan tingkat kemiskinan dalam pemodelan panel dinamis 2013–2022 (Samiani, Endang, Susilo, & Astuti, 2024).

Penyusunan kebijakan yang tepat sasaran merupakan hal penting dalam upaya pemberantasan kemiskinan (Murdiyana & Mulyana, 2017; Sarjito, 2023). Oleh karena itu pemerintah membutuhkan pemetaan wilayah agar upaya pemberantasan kemiskinan ini dapat lebih efektif dan optimal. Salah satu pendekatan untuk melakukan pengelompokan wilayah dengan metode clustering, yakni teknik analisis yang dapat mengelompokkan data berdasarkan kemiripannya (Fathia et al., 2016). Algoritma K-Means merupakan salah satu metode clustering yang umum digunakan, Namun algoritma ini memiliki kelemahan yakni sensitif terhadap pemilihan titik pusat awal (centroid) , di mana hal tersebut berpengaruh terhadap stabilitas hasil cluster (Arthur & Vassilvitskii, 2007).

Penelitian mengenai kemiskinan yang menerapkan algoritma K-Means telah banyak dilakukan. Seperti yang dilakukan Ferdiansyah et al (2024) yang menerapkan algoritma K-Means dalam mengelompokkan wilayah di Jawa Timur tahun 2020 berdasarkan indikator kemiskinan. Namun, indikator yang digunakan tidak dijelaskan secara gamblang serta algoritma yang digunakan terbatas pada algoritma K-Means tanpa membandingkan dengan algoritma yang lain berpotensi memberikan hasil cluster yang lebih akurat.

Penelitian ini mengusulkan pengelompokan wilayah di Jawa Timur tahun 2024 dengan indikator penduduk miskin, TPT, HLS, RLS, TPAK, dan Gini Ratio serta membandingkan hasil pengelompokkan menggunakan algoritma K-Means dan K-Means++. Algoritma K-Means++ yang dikembangkan oleh Arthur & Vassilvitskii (2007) sebagai perbaikan dari algoritma K-Means yang mana cara kerja dalam menginisialisasi centroid dengan didasarkan pada probabilitas yang proporsional terhadap kuadrat jarak titik data ke centroid terdekat yang telah dipilih sebelumnya.

Penelitian ini bertujuan untuk mengetahui karakteristik kabupaten/kota di Provinsi Jawa Timur pada tahun 2024 berdasarkan indikator jumlah penduduk miskin, Tingkat Pengangguran Terbuka (TPT), Harapan Lama Sekolah (HLS), Rata-Rata Lama Sekolah (RLS), Tingkat Partisipasi Angkatan Kerja (TPAK), dan Gini Ratio, serta menganalisis hasil pengelompokan daerah menggunakan algoritma K-Means dan K-Means++. Dengan memahami karakteristik dan pengelompokan ini, penelitian diharapkan memberikan manfaat berupa identifikasi daerah-daerah prioritas yang memerlukan intervensi kebijakan pengentasan kemiskinan secara tepat sasaran,

sehingga pemerintah dan pemangku kepentingan dapat merancang program pembangunan sosial dan ekonomi yang lebih efektif, efisien, dan berbasis bukti, serta meningkatkan kesejahteraan masyarakat secara merata di seluruh kabupaten/kota di Jawa Timur.

METODE

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur tahun 2024, sebagaimana dipublikasikan dalam Indeks Pembangunan Manusia Provinsi Jawa Timur 2020 Volume 7, dengan unit penelitian 38 kabupaten/kota di Provinsi Jawa Timur. Variabel yang digunakan meliputi jumlah penduduk miskin (jiwa), tingkat pengangguran terbuka (persen), harapan lama sekolah (tahun), rata-rata lama sekolah (tahun), tingkat partisipasi angkatan kerja (persen), dan Gini Ratio. Data dikumpulkan melalui publikasi resmi BPS dan diolah menggunakan metode klasterisasi berbasis algoritma K-Means dan K-Means++, yang bertujuan untuk mengelompokkan kabupaten/kota berdasarkan karakteristik sosial-ekonomi tersebut. Analisis data dilakukan dengan membandingkan hasil pengelompokan dari kedua algoritma untuk mengetahui konsistensi dan akurasi klasifikasi, sehingga penelitian ini dapat memberikan manfaat berupa identifikasi daerah-daerah prioritas yang membutuhkan kebijakan pengentasan kemiskinan yang tepat sasaran serta mendukung perencanaan pembangunan sosial-ekonomi yang berbasis bukti di Jawa Timur. Daftar varibel disajikan pada Tabel 1.

Tabel 1. Varibel Penelitian

| No | Variabel | Satuan |
|-----------|------------------------------------|---------------|
| 1 | Jumlah Penduduk Miskin | Jiwa |
| 2 | Tingkat Pengangguran Terbuka | Persen |
| 3 | Harapan Lama Sekolah | Tahun |
| 4 | Rata-Rata Lama Sekolah | Tahun |
| 5 | Tingkat Partisipasi Angkatan Kerja | Persen |
| 6 | Gini Ratio | |

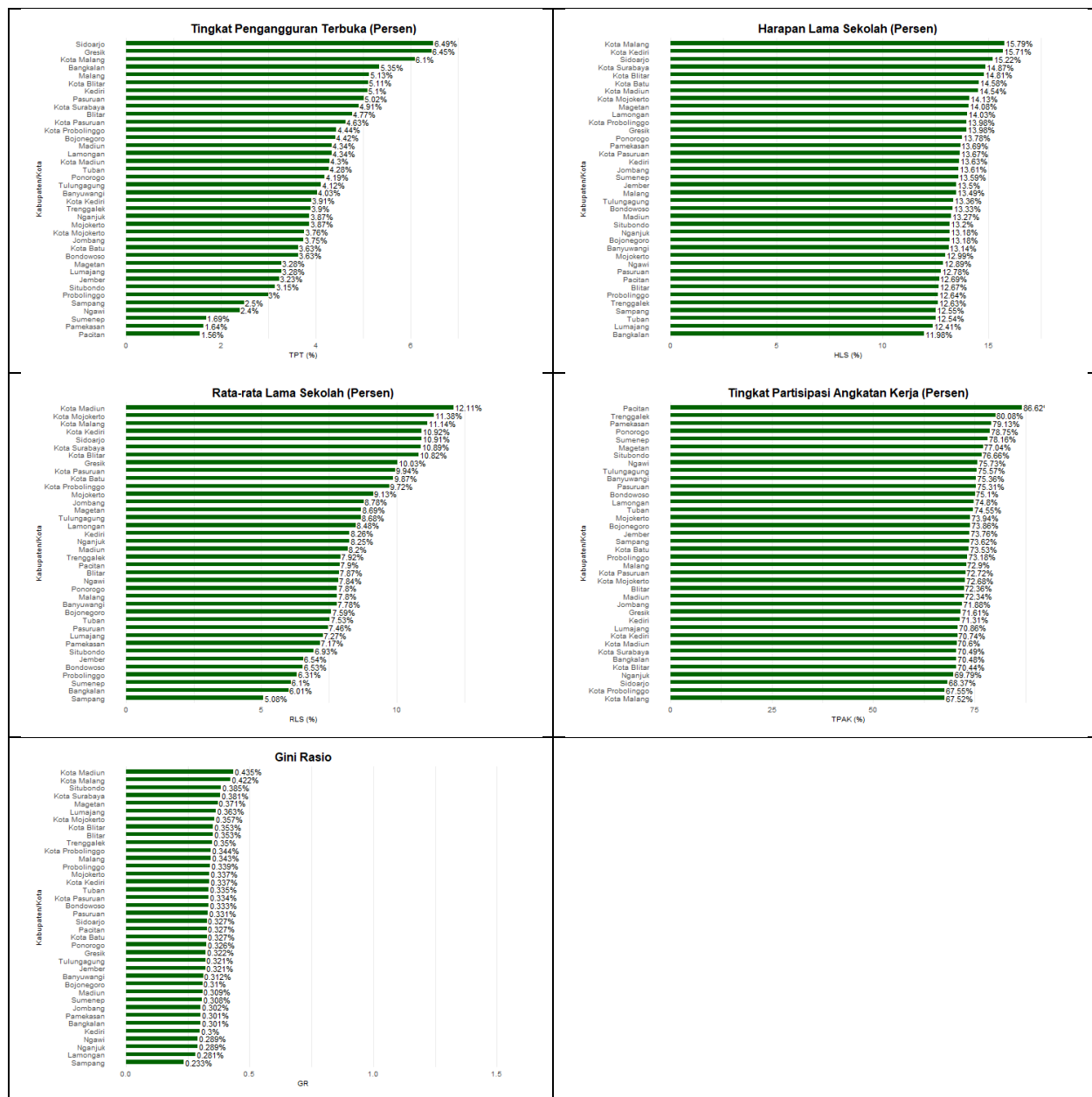
Badan Pusat Statistik, 2025

HASIL DAN PEMBAHASAN

Karakteristik Variabel

Karakteristik masing-masing variabel sebagai indikator untuk clusterisasi kemiskinan yaitu, variabel tingkat pengangguran terbuka, harapan lama sekolah, rata-rata lama sekolah, tingkat partisipasi Angkatan kerja, dan gini rasio ditunjukkan pada Gambar 1. Sebagai berikut.

Pengelompokan Kemiskinan Di Provinsi Jawa Timur Tahun 2024 Dengan Pendekatan Algoritma K-Means Dan K-Means++



Gambar 1. Karakteristik Variabel

Gambar 1 menunjukkan bahwa 15 Kabupaten/Kota di Provinsi Jawa Timur nilai TPT nya masih berada di atas rata-rata TPT Provinsi Jawa Timur pada Tahun 2024 sebesar 4,33%. Kabupaten Sidoarjo menempati posisi tertinggi dengan selisih yang tidak terlalu beda jauh dengan Kabupaten Gresik. Hal ini dapat mengindikasikan adanya potensi tekanan ekonomi yang lebih besar, karena ketika nilai TPT tinggi artinya masih banyak penduduk usia kerja yang belum terserap dalam lapangan pekerjaan. Namun, apabila dilihat dari segi Pendidikan wilayah perkotaan memiliki capaian dan harapan Pendidikan yang lebih tinggi dibandingkan Kabupaten/Kota di wilayah Timur Provinsi Jawa Timur dan Pulau Madura yang beberapa masih berada di bawah rata-rata nasional 13,08% untuk harapan lama sekolah dan 8,27% untuk rata-rata lama sekolah.

Variabel tingkat partisipasi Angkatan kerja, Kabupaten Pacitan memiliki nilai yang paling tinggi, sedangkan Kota Malang, Kabupaten Probolinggi, dan Kabupaten Sidoarjo menempati tiga terbawah. Hal ini dapat disebabkan karena TPT yang masih tinggi dari ketiga kabupaten/Kota tersebut. Variabel selanjutnya adalah gini rasio, di mana apabila nilai gini rasio makin mendekati 0 menunjukkan tidak adanya ketimpangan. Dalam clusterisasi kemiskinan, kombinasi TPAK dan Gini Rasio dapat dipertimbangkan karena apabila nilai TPAK tinggi namun Gini rendah bisa mencerminkan bahwa hampir semua penduduk berada dalam kondisi ekonomi yang setara namun rendah yang mana pada Grafik dapat ditunjukkan oleh Kabupaten Sumenep.

Deteksi Multikolinearitas dan Kaiser Meyer Olkin (KMO)

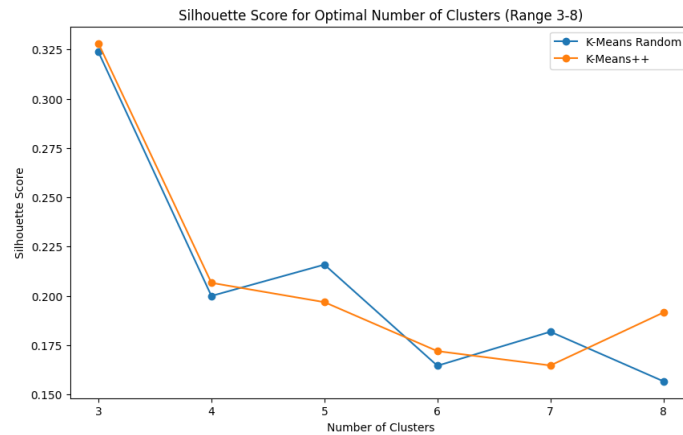
Deteksi multikolinearitas digunakan untuk memastikan bahwa antar variabel tidak terjadi korelasi yang tinggi sehingga masing-masing variabel dapat memberikan informasi yang utuh, pada analisis ini deteksi multikolinearitas dilihat dari nilai pearson correlation. Selain itu dalam analisis cluster juga memperhatikan besarnya nilai KMO yang digunakan untuk memastikan kecukupan data dalam analisis cluster. Pada analisis ini didapatkan nilai KMO sebesar 0,69 yang lebih besar dari 0,5 sehingga data sudah dikatakan cukup untuk digunakan dalam analisis cluster. Diperkuat juga dengan nilai pearson correlation dari masing-masing variabel yang kurang dari 0,8 sehingga tidak terdeteksi adanya multikolinearitas

Hasil Clustering K-Means dan K-Means++

Terdapat dua metode yang dilakukan dalam penelitian ini, yaitu dengan metode K-Means dan K-Means++. Untuk penerapan kedua metode tersebut, terlebih dahulu dilakukan validitas untuk jumlah cluster yang optimum menggunakan nilai DBI dan Silhouette Score dari masing-masing metode yang digunakan.

Table 2. Nilai DBI Hasil Clustering K-Means dan K-Means++

| Jumlah Cluster | Kmeans | Kmeans++ |
|-----------------------|---------------|-----------------|
| 3 | 1,0959 | 1,0786 |
| 4 | 1,3405 | 1,3113 |
| 5 | 1,2614 | 1,3138 |
| 6 | 1,2486 | 1,2217 |
| 7 | 1,2486 | 1,2217 |
| 8 | 1,2623 | 1,1484 |



Gambar 2. Nilai Silhoutte Score Hasil Clustering K-Means dan K-Means++

Pada kedua indeks validitas diatas, untuk menentukan jumlah cluster optimum adalah dengan melihat nilai DBI paling kecil dan Silhoutte Index yang paling besar. Berdasarkan tabel diatas, diketahui bahwa untuk k optimum ada pada jumlah kluster k=3 untuk metode K-Means maupun K-Means++.

Evaluasi Metode Clustering Terbaik

Penentuan metode terbaik dapat dilihat dari nilai inertia masing-masing metode. Nilai inertia untuk metode K-Means dan K-Means++ dapat dilihat pada tabel berikut

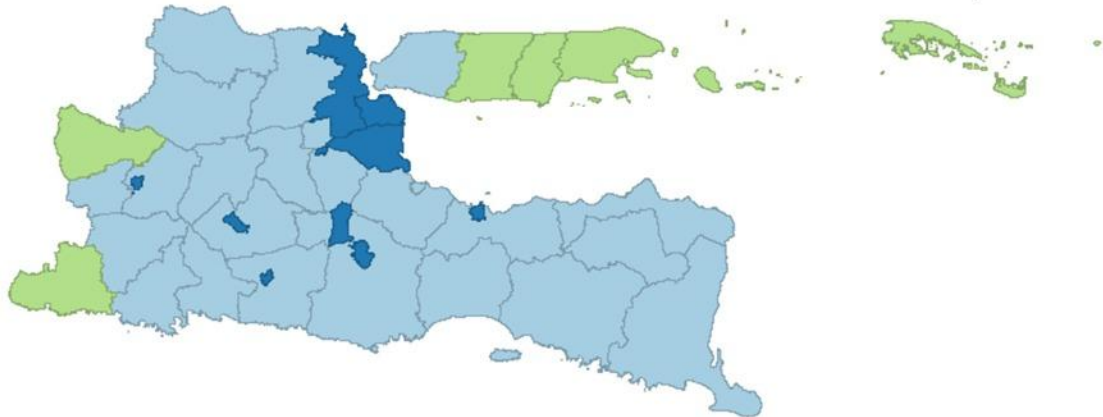
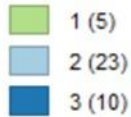
Table 3. Pemilihan Metode Terbaik

| Metode | Jumlah Cluster | Inertia |
|------------------|----------------|---------|
| K-Means | 3 | 87,52 |
| K-Means++ | 3 | 87,37 |

Berdasarkan nilai inertia terkecil dapat disimpulkan bahwa metoded terbaik dalam pengelompokan kemiskinan di Jawa Timur adalah metode K-Means++ dengan jumlah cluster sebesar 3.

Pembahasan

Cluster Kemiskinan Jawa Timur



Gambar 3. Peta Cluster Kabupaten/Kota di Jawa Timur

Berdasarkan hasil analisis pemilihan metode, algoritma K-Means++ lebih baik dalam clusterisasi kemiskinan berdasarkan lima variabel yang digunakan yaitu TPT, HLS, RLS, TPAK, dan GR. Hasil clusterisasi diperoleh tiga cluster optimum dalam pengelompokan kemiskinan di Provinsi Jawa Timur pada Tahun 2023. Hasil cluster center ditunjukkan pada Tabel 3.

Cluster 1 memiliki karakteristik wilayah dengan capaian pendidikan yang relatif tinggi dibanding cluster lainnya. Hal ini dapat dilihat dari nilai HLS dan RLS yang tertinggi dengan masing-masing 14,76 tahun dan 10,78 tahun. Namun, TPT pada cluster 1 masih menunjukkan nilai yang tinggi yaitu sebesar 4,91%. Dari nilai cluster center ini dapat diindikasikan bahwa adanya ketidaksesuaian antara kompetensi lulusan dan kebutuhan lapangan kerja yang akhirnya meningkatkan pengangguran. Apabila hal ini tidak segera ditangani dapat menyebabkan Kab/Kota pada Cluster 1 yang ditunjukkan pada Gambar 1 ditandai dengan warna hijau dengan anggota wilayah terdiri dari 5 wilayah dapat mengalami kemiskinan struktural karena ketidakefisienan dalam pemanfaatan sumber daya manusia.

Cluster 2 terdiri dari 23 wilayah yang ditunjukkan pada Gambar 1 berwarna biru. Berdasarkan nilai cluster centroid yang ditunjukkan pada Tabel 3 dari segi tingkat pendidikan paling rendah dibanding cluster lainnya. Namun, TPAK nya tinggi dan TPT nya rendah. Selain itu dilihat dari Gini Rasionya juga yang paling rendah sebesar 0,2916 menunjukkan adanya tingkat kesejahteraan yang merata namun masih pada level rendah karena sebagian besar penduduk bekerja namun produktivitas rendah sehingga berisiko mengalami kemiskinan fungsional.

Cluster 3 menunjukkan karakteristik Kab/Kota berdasar lima variabel yang digunakan, kelima memiliki nilai yang relatif sedang dibanding dua cluster sebelumnya. Anggota cluster 3 ditunjukkan pada Gambar 1 berwarna biru tua dengan anggota 10 wilayah. Secara keseluruhan Kab/Kota pada cluster ini menunjukkan keseimbangan dari sisi pendidikan, namun angka pengangguran masih termasuk lebih tinggi dibanding cluster 2. Apabila dilihat dari nilai Gini Rasionya tetap harus dioptimalkan terlebih dari peningkatan pendidikan dan penurunan pengangguran agar ketimpangan dalam masyarakat tidak semakin meningkat.

KESIMPULAN

pengelompokan kabupaten/kota berdasarkan variabel jumlah penduduk miskin, TPT, HLS, RLS, TPAK, dan Gini ratio menghasilkan tiga cluster utama dengan karakteristik sebagai berikut. Cluster 1: Menunjukkan indikasi kemiskinan struktural. Wilayah dengan pendidikan tinggi tetapi ketimpangan dan pengangguran juga tinggi. Cluster 2: Menggambarkan kemiskinan fungsional, di mana meskipun masyarakat bekerja, rendahnya tingkat pendidikan dan produktivitas tetap menahan kemajuan ekonomi. Cluster 3: Karakteristiknya cenderung sedang pada seluruh indikator. Algoritma K-Means++ memberikan hasil pengelompokan/clustering yang lebih optimal dibandingkan K-Means, dengan nilai inertia lebih kecil, serta lebih baik berdasarkan nilai DBI dan Silhouette Index. Hasil pengelompokan kluster dapat menjadi bahan pertimbangan dalam merancang strategi untuk penanggulangan kemiskinan yang mana dapat disesuaikan dengan karakteristik masing-masing wilayah agar dapat lebih tepat sasaran.

DAFTAR PUSTAKA

- Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035.
- Badan Pusat Statistik. (2024, July 1). In March 2024, the poor population percentage decreased into 9.03 percent. <https://www.bps.go.id/en/pressrelease/2024/07/01/2370/in-march-2024--the-poor-population-percentage-decreased-into-9-03-percent-.html>
- Badan Pusat Statistik. (2025a). *Indeks pembangunan manusia* (Vol. 7).
- Badan Pusat Statistik Provinsi Jawa Timur. (2025, January 15). The percentage of poor people in Jawa Timur as of September 2024 decreased to 9.56 percent. <https://jatim.bps.go.id/en/pressrelease/2025/01/15/1474/the-percentage-of-poor-people-in-jawa-timur-as-of-september-2024-decreased-to-9-56-percent.html>
- Bappeda Provinsi Jawa Timur. (2025, January 16). Persentase penduduk miskin di Jawa Timur September 2024 turun menjadi 9,56 persen. <https://bappeda.jatimprov.go.id/2025/01/16/persentase-penduduk-miskin-di-jawa-timur-september-2024-turun-menjadi-9-koma-56-persen/>
- Chamidah, N., Mardianto, M. F. F., Limanta, E. E., & Hastuti, D. R. (2020). Modelling of poverty percentage based on mean years of schooling in Indonesia using local linear estimator. In *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)* (pp. 87–91). Atlantis Press. <https://doi.org/10.2991/assehr.k.201010.014>
- Fathia, A. N., Rahmawati, R., & Tarno. (2016). Analisis kluster kecamatan di Kabupaten Semarang berdasarkan potensi desa menggunakan metode Ward dan single linkage. *Jurnal*

Gaussian, 5, 801–810.

- Faturohim, A., Akbar, A., Hidayat, B. A., & Saksono, H. (2023). An analysis of urban poverty and unemployment. *Jurnal Bina Praja*, 15(2), 309–324. <https://doi.org/10.21787/jbp.15.2023.309-324>
- Ferdiansyah, I., Huda, B., Hananto, A., & Tukino. (2024). Analisis clustering menggunakan metode K-means pada kemiskinan di Jawa Timur tahun 2020. *Journal of Social Science Research*, 4, 858–869.
- Murdiyana, M., & Mulyana, M. (2017). Analisis Kebijakan Pengentasan Kemiskinan Di Indonesia. *Jurnal Politik Pemerintahan Dharma Praja*, 73–96.
- Purwono, R., Wardana, W. W., Haryanto, T., & Mubin, M. K. (2021). Poverty dynamics in Indonesia: Empirical evidence from three main approaches. *World Development Perspectives*, 23, 100346. <https://doi.org/10.1016/j.wdp.2021.100346>
- Samiani, S., Endang, E., Susilo, J. H., & Astuti, H. (2024). Dynamic panel data modeling of Indonesia's poverty level 2013–2022. *Jurnal Ekonomi & Studi Pembangunan*, 25(1), 130–147. <https://doi.org/10.18196/jesp.v25i1.21079>
- Sarjito, A. (2023). Efektivitas Kebijakan Sosial dalam Mengurangi Ketimpangan Pendapatan dan Angka Kemiskinan. *Jurnal Ilmu Sosial Politik Dan Humaniora*, 6(2), 1–12.
- Suryahadi, A., Izzati, R. A., & Suryadarma, D. (2020). Estimating the impact of COVID-19 on poverty in Indonesia. *Bulletin of Indonesian Economic Studies*, 56(2), 175–192. <https://doi.org/10.1080/00074918.2020.1779390>
- Taufiq, N., & Dartanto, T. (2020). Education, informal turnover and poverty dynamics in Indonesia. *International Journal of Economics and Management*, 14(1), 157–172.
- United Nations Development Programme. (2023). *Multidimensional Poverty Index (MPI) 2023: Indonesia country profile*. <https://hdr.undp.org/sites/default/files/Country-Profiles/MPI/IDN.pdf>
- World Bank. (2023, May 9). *Indonesia poverty assessment: Pathways towards economic security*.